普通高等教育“十一五”国家级规划教材

医学信息检索教程

（第2版）

Medical Information Retrieval

董建成　主编

东南大学出版社
SOUTHEAST UNIVERSITY PRESS

普通高等教育“十一五”国家级规划教材

医学信息检索教程

第2版

主　编　董建成

副主编　蒋　葵　张志美　蔡忆宁

编　者　（按姓氏笔画排序）

马　路（首都医科大学）

王秀平（山西医科大学）

王晓梅（江苏省科技情报所）

叶春峰（西安交通大学）

李爱国（东南大学）

吴华珠（江苏省科技情报所）

张志美（南通大学）

张燕蕾（北京大学）

胡小君（浙江大学）

胡新平（南通大学）

顾　骏（南通大学）

梅　谊（苏州大学）

符礼平（复旦大学）

谢志耘（北京大学）

蒋　葵（南通大学）

董建成（南通大学）

蔡忆宁（江苏省科技情报所）

东南大学出版社

南　京

内容提要

作为普通高等教育“十一五”国家级规划教材，《医学信息检索教程》第2版及时反映了医学信息资源和信息技术的最新进展。本书在阐述信息类型、检索语言、检索途径、检索技术、网络基础和数字图书馆等基本知识的基础上，详细介绍了各类医学信息资源及其检索方法。内容包括网络检索工具，如Google、Yahoo、Medical Matrix、Medscape等；经典的生物医学数据库，如CBMDisc、MedLine、PubMed、EMBASE、BIOSIS和IPA等；综合性文献数据库资源，如CNKI、万方数据资源、NSTL、超星数字图书馆、SDOL、EBSCO、Springer等；免费医学信息资源，如Free Medical Journals、HighWire Press、PMC等；特定类型的医学信息，如循证医学信息、引文信息、特种文献信息和医学参考工具书。为了进一步提高读者对医学信息资源的分析和利用能力，本书将科技查新和信息分析的有关知识纳入其中。

本书图文并茂，内容详尽，将枯燥的信息检索语言融于生动的数据库实践操作中，有利于激发读者的学习兴趣，有利于提高学习质量，便于读者自学，适合于医学高等院校的本科生和研究生、医务工作者及医药信息学相关人员作为教科书和参考书。

图书在版编目(CIP)数据

医学信息检索教程 / 董建成主编. —2版. —南京：东南大学出版社，2009.2

ISBN 978-7-5641-1542-5

Ⅰ.医… Ⅱ.董… Ⅲ.医药学—情报检索—高等学校—教材 Ⅳ.G252.7

中国版本图书馆CIP数据核字(2009)第001987号

东南大学出版社出版发行

(南京四牌楼2号 邮编210096)

出版人：江 汉

江苏省新华书店经销 南京京新印刷厂印制

开本：787mm×1092mm 1/16 印张：27.5 字数：686千字

2009年2月第2版 2009年2月第10次印刷

ISBN 978-7-5641-1542-5/R·126

印数：29001～33000册 定价：45.00元

(凡因印装质量问题，请直接向出版社读者服务部调换。电话：025－83792328)

序

信息素养(Information Literacy)是一个人能够认识到何时需要信息,有效地检索、评价和利用信息的能力。信息素养可为一生学习奠定基础,适用于各个学科、各种学习环境和教育水平。信息素养的培养目标是使学习者能够决定所需信息的范围,有效地获取信息,准确地评价信息,将所选信息融合到个人的知识体系之中,并有效地运用信息达到特定目的。

随着信息化社会进程的加快,来自图书馆、互联网、社区、媒体等越来越多的未经过滤的信息使得它们失去了真实性、正确性和可靠性,人们很难获取和评价以图片、声像和文本等形式存在的信息。因此,信息素养在当代科技迅速发展和信息资源极其丰富的环境下变得越来越重要。

由董建成教授主编的《医学信息检索教程》(第2版)是针对当前医学院校的本科生、研究生教育和临床医生继续教育的需求,更为全面地分析、研究和介绍了医学信息检索的基本原理和方法。其内容不但包含了信息检索基础知识、网络信息检索工具、中外文医学信息检索、循证医学信息检索、引文信息检索,还介绍了特种文献检索、医学参考工具书利用、科技查新、信息分析与研究等,构成了医学信息检索与利用课程的有机整体。教材的参编者多为国内长期从事医学信息检索教学和科研的资深专家,汇集了他们多年的医学信息检索教学和研究的经验,整个教程的结构严谨,层次清晰,内容丰富,重点突出,通俗易懂,实用性强,具有颇高的教学使用和参考价值,不仅适合于高等医学院校的本科生、研究生学习和参考,亦可作为医学研究人员、临床医务人员和医学信息相关人员的工具书和参考书。

医学信息检索课程是一门实践性强、应用性广、内容更新快的科学方法课程,是医科大学生和广大医学科研人员信息素养教育的重要环节。该教材在第一版的基础上,结合现代信息技术和学科发展的需要,从整体上作了相应的调整、充实和完善,是全体作者从事医学信息检索教学和科研的新成果与经验总结,是站在医学信息学的学科发展高度总结和研究了医学信息检索的原理、方法和技术,这也正是本书的可贵之处。希望高等医学院校的各类学生、医务工作者和广大读者认真研读,细细品味,定能起到事半功倍的效果。

全国医学信息检索教学研究会会长
复　旦　大　学　教　授　徐孝勤

2008年12月于复旦大学

前　言

《医学信息检索教程》第1版自2002年8月出版以来，先后被上海、江苏、浙江、安徽、山东、甘肃等省市的高校选作教材，在培养高等医学院校学生和医务工作者信息素养的过程中发挥了一定的作用。为构建《医学信息检索》课程的立体教材体系，编写组又完成了《医学信息检索》CAI课件的编写和制作，并由人民卫生电子音像出版社出版，与《医学信息检索教程》共同获得了江苏省高等教育教学成果二等奖。

社会和科学在不断地发展，信息技术在不断地进步，信息资源在不断地更新和增长。编写组成员广泛听取了国内高校任课教师和学生对《医学信息检索》课程教学的意见和建议，在认真总结第1版编写工作的基础上，提出了修订计划，及时反映信息技术和信息资源的最新进展，编写了《医学信息检索教程》第2版，以飨读者。本教材已被列入普通高等教育“十一五”国家级规划教材。

《医学信息检索教程》第2版共分十章，内容包括四部分：第一部分（第一至二章），简要介绍医学信息检索的基础理论和基本概念，包括医学信息的类型，信息检索的语言、方法、途径、步骤和技术等，并对与信息检索密切相关的计算机网络、数字图书馆和网络检索工具进行了介绍。第二部分（第三至四章），以中国生物医学文献数据库、中国知识基础设施工程、万方数据知识服务平台、维普中文科技期刊数据库、国家科技图书文献中心和超星数字图书馆为例，介绍了中文医学信息检索的方法和途径；以PubMed、EMBASE、BIOSIS、SciFinder、IPA、CC等为例，介绍了国外主要生物医学数据库的最新检索技术；并向读者推荐了SDOL、EBSCO、Springer、UMI ProQuest、CSA、OCLC FirstSearch等综合性外文数据库和Free Medical Journals、HighWire Press、PMC等免费信息资源。第三部分（第五至八章），着眼于专类医学信息，包括循证医学信息、引文信息和特种文献信息的检索，介绍了利用医学参考工具书进行事实检索和数据检索的方法。第四部分（第九至十章），主要介绍科技查新和信息分析的有关知识，为医学生和医务工作者更好地利用医学信息资源提供帮助。

本教材汲取了国内外许多专家学者的有关研究成果，并承蒙全国医学信息检索教学研究会会长徐一新教授为本教材作序，在此一并致谢。

限于水平，书中难免有欠妥之处，殷请广大师生和读者不吝赐教，惠予指正。

2008年11月于南通大学

目　录

第一章　信息检索基础

信息检索(Information Retrieval)是指信息的有序化识别和查找的过程，即人们根据特定的信息需求，采取科学的方法，应用专门的工具，从浩瀚的信息海洋中迅速、准确地获取所需信息的过程。

早期的信息检索，人们主要根据文献的特征，用手工方式实现。以计算机为核心的信息技术，开辟了信息处理与信息检索的新纪元，计算机从处理数字信息发展到处理字符信息、静态和动态的图像信息乃至声音信息等，不仅拓展了信息检索的领域，丰富了信息检索的内容，而且极大地提高了信息检索的速度。近年来，互联网的普及，给检索工作带来了一个全新的发展空间，信息检索的对象已从过去相对封闭，由独立数据库集中管理的信息内容扩展到如今开放、动态、更新更快、分布广泛、管理松散的网络内容；网络信息检索从一开始的一般人难以学会的标准化检索发展到现在，已经成为简单的、大众化的行为方式了。信息检索已成为当今科学研究、经济活动和社会生活中的一个组成部分并发挥越来越大的作用。

第一节　信息与信息资源

一、信息的涵义

信息是许多学科广泛使用的概念，在不同的学科领域有着不同角度的解释。但人们普遍认为信息与能源、材料科学并列，构成现代社会的三大支柱。

在信息检索领域，一般将信息理解为关于现实世界事物存在方式或运动状态的反映。例如，作为医疗对象的某病人，年龄58岁，性别男，身高1.72 m，体重69 kg，体温37.2℃，患有糖尿病，这些都是关于某病人的信息，是某病人存在状态的反映。

信息有许多重要的特征：信息来源于物质和能量；信息是可以感知的；信息是可以存储的；信息是可以加工、传递和再生的。这些特征构成了信息的最重要的自然属性。作为信息的社会属性，信息已经成为各行各业不可缺少的重要资源之一。人类获取、积累并利用信息是认识和改造客观世界的必要过程。借助信息，人类才能获得知识，才能有效地组织各种社会活动。因此，信息是人类维持正常活动不可缺少的资源。

二、信息的类型

1. 文字信息

文字是人们为了实现信息交流、通信联系所创造的一种约定的形象符号。广义的文字

还包括各种编码，如ASCII码、汉字双字节代码、国际电报与单元代码以及计算机中的二进制数字编码等。

2. 图像信息

图像是一种视觉信息，它比文字信息直接，易于理解。人工创造的图像，如一张纸、一幅画、一部电影，大自然的客观景象等都是抽象或间接的图像信息。随着多媒体技术的发展，各类图像信息库将会极大地丰富人类生活。

3. 数值数据信息

数值数据是“信息的数字形式”或“数字化的信息形式”。狭义的数据是指有一定数值特性的信息，如统计数据、气象数据、测量数据以及计算机中区别于程序的计算数据。广义的数据是指在计算机网络中存储、处理、传输的二进制数字编码。文字信息、图像信息、语音信息以及从自然界直接采集的各种自然信息均可转换为二进制数码，网络中的数据通信、数据处理和数据库等就是广义的数值数据信息。

4. 语音信息

讲话，实际上是人大脑中的某种编码形式的信息转换成语音信息的输出，是一种最普遍的信息表现形式。音乐也是一种信息形式，是一种特殊的声音信息，它是通过演奏方式来表达丰富多彩的信息内容的。

三、信息资源的涵义

信息资源是人类在认识世界与改造世界过程中所产生、整理和记录的有用信息的集合。信息资源是信息与资源两个概念整合衍生出来的新概念，它归根结底是一种信息，或者说是信息的子集。而资源是通过人类的参与而获取的(或可获取的)可利用的物质、能量与信息的总和。联系信息概念与资源概念来考察信息资源，可以这样认为：①信息资源是信息的一部分，是信息世界中与人类需求相关的信息；②信息资源是可利用的信息，是在当前生产力水平和研究水平下人类所开发与组织的信息；③信息资源是通过人类的参与而获取的信息。人类的参与在信息资源形成过程中具有重要的作用。总之，信息资源就是经过人类开发与组织的信息的集合，而“开发与组织”正是信息资源可利用的表征。

四、信息资源的类型

信息资源的类型可以根据多种标准来划分。

以开发程度为依据，信息资源可划分为潜在的信息资源与现实的信息资源两大类。潜在的信息资源是指个人在认知创造过程中储存在大脑中的信息资源，它们虽能为个人所利用，但一方面易于随忘却过程而消失，另一方面又无法为他人直接利用，因此是一种有限再生的信息资源。现实的信息资源则是指潜在信息资源经个人表述之后能够为他人所利用的信息资源，他们最主要的特点是具有社会性，通过特定的符号表述和传递，可以在特定的社会条件下广泛地连续往复地为人类所利用，因此是一种无限再生的信息资源。

现实信息资源以表述方式为依据可以划分为口语信息资源、体语信息资源、文献信息资源和实物信息资源。

口语信息资源是人类以口头语言所表述出来而未被记录下来的信息资源，它们在特定的场合被直接消费并且能够辗转相传而为更多的人所利用，如谈话、聊天、授课、讲演、讨论、

歌唱等活动都是以口语信息资源的交流和利用为核心的。

体语信息资源是人类以手势、表情、姿势等方式表述出来的信息资源。它们通常依附于特定的文化背景，如舞蹈就是一种典型的体语信息资源。

实物信息资源是人类通过创造性的劳动以实物的形式表述出来的信息资源。这类信息资源中物质成分较多，有时难以区别于物质资源，而且它们的可传递性一般较差。实物信息资源有产品样本、模型、碑刻、雕塑等。

文献信息资源是以语言、文字、图像、声频、视频等方式记录在特定载体上的信息资源，最主要的特征是拥有不依附于人的物质载体，只要这些载体不损坏或消失，文献信息资源就可以跨越时空无限往复地为人类所利用。

文献信息资源以记录方式和载体的形式为依据可划分为印刷型、缩微型、声像型、电子型等。

1. 印刷型

印刷型文献又称纸质型文献，是指以手写或印刷技术为主要手段、以纸张为信息记录载体的文献。其优点是可以直接阅读，携带方便，是目前人类信息交流活动中最常用的工具。与现代信息载体相比，印刷型文献存储信息密度低，占用收藏空间大，不易长期保存，难以实现自动化输入和自动检索。印刷型文献可分为三类：

(1) 图书。图书(Book)通常提供比较系统、成熟的知识，一般包括专著、教科书、丛书、论文集和参考工具书等。专著是对某一个专题有较深入的研究和独到见解的学术著作，如《心血管药理学》、《休克》等。教科书是某个专业或学科的研究总结，反映较成熟的专业理论，具有严格的系统性与逻辑性，内容可靠性强，是医学生和医学工作者进行专业学习的主要医学文献。论文集是由多位作者的论文或会议论文、报告等汇编而成的出版物。参考工具书是供日常工作、学习或写作中随时查阅用的一类图书，其内容有序，便于查考，主要包括字典、词典、年鉴、手册、名录、图谱、百科全书等。

(2) 期刊。期刊(Journal)也叫杂志，是指具有相对固定的刊名、编辑机构及版式装帧的连续出版物，如美国的《Science》(科学)、英国的《Nature》(自然)、我国的《中华医学杂志》等。期刊的内容通常是能够反映学科领域最新的理论、方法、技术的论文(Journal Article)、综述(Review)、病例报告(Case Report)等。期刊论文包括研究报告、论著、著述等，是反映最新科研成果，具有学术性、创新性和科学性特点的信息。综述是综合描述某一专题或学科在一定时间内的研究现状和进展的文献，其综合性强，权威性高，能够直接反映专业领域内科研的动向和进展。

(3) 特种文献。特种文献又称非书非刊资料，包括除图书、期刊以外的其他出版物，常为不定期出版，多数具有连续性。其特点是：数量大、种类多、内容广、参考价值大。

①政府出版物：是指国家各级政府部门及其所属机构出版的文献信息资料，主要包括社会科学与自然科学两大类。其中，行政文件，如讨论会记录、各种法令、外交文件等，一般统计数据占多数，科技资料相对较少。

②会议文献：是指在国内外学术团体举行的专业会议上发表的论文或学术报告，其特点是信息传播速度快，反映研究成果新。会议文献主要通过会议论文摘要、论文集、期刊特辑或增刊等形式予以刊载。

③专利文献：专利(Patent)是指受到法律保护的技术发明。专利文献是指发明人向政

府部门(专利局)递交的、说明自己创造的技术文件,同时也是实现发明所有权的法律性文件,包括专利说明书、专利公报、商标等,具有新颖性、创造性、实用性等特征。

④科技报告:是指各学术团体、科研机构、高等院校的研究报告及其研究过程的记录,其理论性较强,是反映某一专业领域科研进展和动态的重要信息。但科技报告保密性强,通常难以获取。

⑤技术标准和规范:又称标准文献,是有关产品或工程质量、规格、生产过程、检验方法的技术文件,具有一定的法律约束力。主要包括技术标准、技术规范、操作规程、准则、术语等。

⑥学位论文:是指高等院校的博士或硕士研究生攻读学位而撰写的毕业论文。

⑦其他:如报纸、手稿、内部刊物、病历档案、技术资料、产品样本,等等。

2. 缩微型

缩微型信息载体是指以感光材料记录信息的载体,如缩微胶卷、缩微胶片、计算机存取载体的输出胶片(Computer Output Microfilm, COM)等。缩微型信息载体体积小、存储信息密度高、成本低廉、便于保存是其优点,但使用时必须借助于阅读机或阅读复印机。

3. 声像型

声像型信息载体又称视听型信息载体,是指记录声音、图像信息的载体,如照片、录音带、录像带、幻灯片、影视片、视听光盘等。声像型信息载体可以让人们通过自己的视觉、听觉感受到直观、形象、生动、逼真、丰富多彩的信息世界。

4. 电子型

电子型文献也称机读型文献、数字型文献,是采用电子手段并以数字形式存储、利用计算机及现代通讯方式提供信息的一种新型信息载体,如光盘数据库、网络数据库、电子图书、电子杂志、电子地图等。数字型信息载体的问世是信息时代的重要标志,它改变了旧有书刊的物理形态,开辟了一种新的信息传播渠道,极大地提高了信息的传递速度,加速了社会信息化的进程。与传统信息载体相比,其优点是信息容量大,传递速度快,便于检索且效率高。电子型文献与印刷型文献共同成为当前科学信息的两大主流载体。常见的电子型文献有以下几种:

(1) 数据库。数据库(Database,DB)可以直观地理解为存放数据的仓库,只不过这个仓库是在计算机的大容量存储器上,如磁盘数据库、光盘数据库(CD-ROM)、联机数据库、网络数据库等。数据库中的数据可以是数字,也可以是文字、图形、图像、声音等,虽然有多种表现形式,但它们都是经过数字化后存入计算机的。

(2) 网络文献。网络文献的出版、传递、检索和利用是通过 Internet 得以实现的。通常利用 WWW(信息浏览)、FTP(文件传输)、Telnet(远程登录)、Gopher(信息查找)、Archie(文件名查询)、USENET(网络新闻)、E-mail(电子邮件)等方式检索 Internet 上的各种各样的信息。

(3) 印刷型文献的数字化。印刷型文献的数字化主要是将印刷型文献数字化后,制成可供计算机阅读、检索和利用的电子出版物,主要有电子图书、电子杂志、电子地图等等。

另外,有学者按对信息加工深度的不同将文献划分为一次文献、二次文献和三次文献。一次文献即原始文献,是作者以生产或科研成果为依据而创作的原始文献,如专著、期刊论文、研究报告、学位论文、发明专利等。二次文献是根据一次文献的内容和外表特征进行加

工整序后的文献，如目录、索引、文摘、书目数据库、搜索引擎等，常被视为信息检索工具的主体。三次文献是对一次和二次文献进行综合、分析后编辑而成的文献，如综述、评论、科技动态、进展、指南等。

五、信息资源的特征

信息资源是可利用的信息，它具有除“无限性”之外信息的所有性质。相对于其他非资源型信息，信息资源具有以下 4 个明显的特征。

1. 智能性

信息资源是人类所开发与组织的信息，是人类脑力劳动或者说认知过程的产物。人类的智能决定着特定时期或特定个人的信息资源的量与质，智能性也可以说是信息资源的“丰度与凝聚度”的集中体现。信息资源的智能性要求人类必须将自身素质的提高和智力开发放在第一位，必须确立教育和科研的优先地位。

2. 有限性

信息资源只是信息的极有限的一部分，比之人类的信息需求，它永远是有限的。从某种意义上说，信息资源的有限性是由人类智能的有限性决定的。有限性要求人类必须从全局出发合理布局和共同利用信息资源，最大限度地实现资源共享，从而促进人类与社会的发展。

3. 不均衡性

由于人们的认识能力、知识储备和信息环境等多方面的条件不尽相同，他们所掌握的信息资源也多寡不等；同时，由于社会发展程度不同，对信息资源的开发程度不同，地球上不同区域信息资源的分布也不均衡，通常所谓的信息领域的“马太效应”就是与这种不均衡性有关的现象。不均衡性要求有关信息政策、法律和规划等必须考虑导向性、公平问题和有效利用问题。

4. 整体性

信息资源作为整体是对一个国家、一个地区或一个组织的政治、经济、文化、技术等的全面反映。整体性要求对所有的信息资源和信息资源管理机构实行集中统一的管理，从而避免人为的分割所造成的资源的重复和浪费。

第二节　信息检索

广义的信息检索包括信息的存储和信息的检索，往往又称为“信息存储与检索”(Information Storage and Retrieval)。信息的存储主要是在一定专业范围内的信息选择基础上进行信息特征描述、加工并使其有序化，或建立数据库，以便在检索时借助一定的设备与工具，从中查找出所需的信息。存储是检索的基础，检索是存储的反过程。在现代信息技术的条件下，信息检索从本质上讲，是指人们从任何信息系统中高效、准确地查找到自己所需的有用信息，而不管它以何种形式出现，或借助于什么样的媒体，此即狭义的信息检索。本书所讲的信息检索主要指的是后者。

一、信息检索系统

信息检索系统是根据社会发展需要和为达到特定的信息交流目的而建立的一种有序化的信息资源集合体。它通常是一个拥有选择、整理、加工、存储、检索信息的设备与方法，并能够向用户提供信息服务的多功能开放系统。信息检索系统一般由下列要素构成：

1. 信息资源

信息资源是系统存储与检索的对象。它可以是全文信息，也可以是题录、索引或文摘；它可以是文字信息，也可以是图形、图像、数值数据或语音信息。

2. 设备

所谓设备，即实现信息存储与检索活动的一切设备。如手工检索的卡片、印刷型检索工具、计算机、交换机、服务器、通讯网络、软件，等等。

3. 方法与策略

包括检索语言、标引方法、信息的组织与管理方法、信息的检索策略与技巧等。

4. 人

人是检索系统的能动因素。充当信息与用户媒介的检索人员，将随着社会网络化程度的不断提高而逐步退出系统，由具有自主检索能力的最终用户取而代之。

二、信息检索类型

1. 按信息检索的对象划分信息检索类型

(1) 文献检索。文献检索就是从大量的文献集合中查找出符合特定需要的相关文献的过程。一般是先查找出相关文献的线索，如题录、文摘等，然后进一步查寻原始文献进行阅读参考。文献检索的结果是有关某课题或特定需要的一组相关性文献。

(2) 数据检索。数据检索是以特定的数值型数据为检索对象的检索过程。包括各种统计数字、图表、化学结构式、计算公式等等，如胰岛素的理化常数、结构式、常用剂量等。

(3) 事实检索。事实检索是利用特定的参考工具书或事实型数据库查找出能够直接解答某一提问的事实。例如，什么是基因工程，何谓生物芯片，何人何时在何处首先提出了人类基因组计划，等等。

综上所述，数据检索、事实检索是一种确定性检索，其检索结果可以直接回答有或无，正确或错误；而文献检索是一种相关性检索，其检索结果只提供与之相关的文献以供参考，不直接回答用户提出的问题。文献检索是信息检索的一个重要组成部分，科技人员在进行信息检索的过程中，通常以文献检索为主。

2. 按信息组织的方式划分信息检索类型

(1) 目录检索。目录检索是指通过卡片式目录、书本式目录、机读目录(Machine Readable Catalog，MARC)或联机公共检索目录(Online Public Access Catalog)查询单位出版物(如一本书、一种杂志、一件专利)的名称、著者、出版事项等文献外表特征的过程，供人们了解出版或收藏机构是否拥有所需的图书、期刊等出版物的情况。

(2) 题录检索。题录检索类似于目录检索，但其检索结果不是单位出版物，而是单位出版物中单篇文献的外表特征。如美国的《医学索引》(Index Medicus，IM)、《中国生物医学文献数据库》(Chinese BioMedical Disc，CBMDisc)等。

(3) 文摘检索。文摘检索是在题录检索的基础上，增加了反映文献的主题范围、目的、方

法、结果等内容特征的摘要。有利于引导用户阅读原文，节约阅读时间，确定所获文献与用户需求的相关程度。如 Medline 数据库、美国《生物学文摘》(BA)、荷兰《医学文摘》(EM)等。

(4) 全文检索。全文检索是用户根据特定的需要，从存储有整篇文章乃至整本图书的信息检索系统中获取全文或有关章节信息的过程。利用全文信息检索系统，还可以进行各种信息的频率统计和内容分析。随着计算机容量与运行速度的不断增大和提高，全文检索正迅速由最初的法律、文学领域扩大到几乎所有的学科和专业领域。

(5) 超文本检索。超文本的基本组成元素是节点(Nodes)和节点间的逻辑联接链(Links)，每个节点中所存储的信息以及信息链被联系在一起，构成相互交叉的信息网络。与传统文本的线性顺序不同，超文本检索强调中心节点之间的语义联接结构，依靠系统提供的复杂工具作图示穿行和节点展示，提供浏览式查询。其检索模式是从“哪里”到“什么”。而传统的文本检索系统则强调文本节点的相对自主性，其检索模式是从“什么”到“哪里”。

(6) 超媒体检索。超媒体检索是对超文本检索的补充。其存储对象超出了文本范畴，融入了静态或动态的图形、图像、声音等多种媒体信息。信息的存储结构从单维发展到多维，存储空间亦在不断扩大。

三、信息检索语言

信息检索语言是为建立信息检索系统而创建的专门用来描述文献特征(内容特征或外表特征)和表达检索提问的一种人工语言，又称为信息存储与检索语言、标引语言、索引语言等。它的主要功能是：①简单明了而又较为专指地描述信息的主题概念；②容易地将概念进行系统排列；③便于检索时将标引用语与检索用语进行相符性比较。因此，信息检索语言不仅需要排除一词多义、多词一义和词义含糊的现象，而且需要显示出概念间的相互关系，这也是信息检索语言规范化的主要内容。

信息检索语言是决定检索系统中大量信息排检序列的关键。它可以是一系列概括信息内容的概念及其相互关系的标识系统，如分类号码；也可以是自然语言中选择出来并加以规范化的一套词汇，如主题词表。世界上有许多种信息检索语言，人们常用的有以下两种。

1. 分类检索语言

分类检索语言是以学科分类为基础，结合信息内容特征的一种直接体现知识分类概念的检索语言。其采用概念逻辑分类的一般规则进行层层划分，构成具有上位类和下位类之间隶属关系、同位类之间并列关系的概念等级体系。例如：

R5 内科学
　R51　传染病
　R52　结核病
　R53　寄生虫病
　R54　心脏、血管(循环系统)疾病
　　R541　心脏疾病
　　.1　先天性心脏血管病
　　.2　风湿性心脏病
　　.3　高血压性心脏病
　　.4　冠状动脉(粥样)硬化性心脏病(冠心病)

分类检索语言的“语词”就是它的类目及相应的分类号。分类号主要用于明确各类目之间的先后顺序。如上例的分类号排序是 R5，R51，R52，R53，R54，R541，R541. 1，R541. 2，R541. 3，R541. 4……

分类检索语言既可以用于期刊论文的分类，也可以用于图书等其他文献信息的分类。国内外有多种广泛使用的著名分类检索语言，如美国《国会图书馆图书分类法》(Library of Congress Classification，LC)、《国际十进分类法》(Universal Decimal Classification，UDC)、《杜威十进分类法》(Dewey Decimal Classification and Relative Index，DC 或 DDC)、《中国图书馆分类法》(中图法)。《中国图书馆分类法》是我国使用最普遍的一种分类检索语言。

《中国图书馆分类法》共分 22 个基本大类(见表 1-2-1)，“R 医药、卫生”类下分 17 个二级类目(见表 1-2-2)。

表 1-2-1 《中国图书馆分类法》基本大类

A	马克思主义、列宁主义、毛泽东思想、邓小平理论	N	自然科学总论
B	哲学、宗教	O	数理科学和化学
C	社会科学总论	P	天文学、地球科学
D	政治、法律	Q	生物科学
E	军事	R	医药、卫生
F	经济	S	农业科学
G	文化、科学、教育、体育	T	工业技术
H	语言、文字	U	交通运输
I	文学	V	航空、航天
J	艺术	X	环境科学、安全科学
K	历史、地理	Z	综合性图书

表 1-2-2 “R 医药、卫生”的二级类目

分类号	类 目	分类号	类 目
R1	预防医学、卫生学	R74	神经病学与精神病学
R2	中国医学	R75	皮肤病学与性病学
R3	基础医学	R76	耳鼻咽喉科学
R4	临床医学	R77	眼科学
R5	内科学	R78	口腔科学
R6	外科学	R79	外国民族医学
R71	妇产科学	R8	特种医学
R72	儿科学	R9	药学
R73	肿瘤学		

2. 主题检索语言

主题检索语言是用表达文献主题内容的词语作为标识的信息检索语言。应用较多的是主题词和关键词。

(1) 主题词法。主题词(Subject Heading)又称叙词(Discriptor),是以规范化为基础,以揭示事物对象及其特征为出发点的信息检索语言。最具代表性的主题词法是美国国立医学图书馆(National Library of Medicine,NLM)的《医学主题词表》(Medical Subject Headings,MeSH)。MeSH 是医学领域内使用最多的一种主题检索语言。MeSH 用于标引和揭示医学文献的主题内容,对于提高医学信息检索的准确率具有十分重要的意义。

随着 Internet 的不断发展和人类信息需求的日益增长,人们在日常的信息检索过程中,越来越重视的是事物的概念和语义,而不容易理解数据库系统的特定句法。所以,自然语言的检索更容易为人们所接受。但传统的自然语言检索,由于检索词与著者使用的文本词不统一,容易造成漏检和误检。因此,在计算机信息检索数据库中,出现了检索词自动转换系统、智能检索系统等,以方便用户进行检索。这些系统是将用户输入的概念和语义自动转换成满足相对查全和查准的数据库系统语言进行检索。如美国 NLM 自 1986 年起研究和开发的一体化医学语言系统(Unified Medical Language System,UMLS),就是在 MeSH 基础上,应用先进的计算机信息技术建立的一个全新的生物医学信息检索语言的集成系统和机读信息资源指南系统,可用于跨数据库的词汇转换,具有一定的数据库集成检索功能和自然语言词语转换等智能检索功能。

UMLS 通过将大量的检索词(包括规范词和自由词)累积输入系统中,进行检索词自动转换处理,使用户能够不必考虑检索词的规范性或知识分类属性,不受人工语言和自然语言的束缚与限制,更自由地在电子病案、文献数据库、图像数据库、专家系统等各种信息资源库中检索和获取特定的信息。UMLS 包括 4 个部分:

①超级叙词表(Metathesaurus):有人将其译为元辞典,是 UMLS 的核心部分,在 2001 年版收录了 80 万个概念共 190 万个词汇。这些概念和词汇来自包括 MeSH 在内的 60 多个生物医学词表、分类表、术语表、专家系统等。其目的是要构建一个整合各来源词表中的生物医学概念、术语、词汇及其等级范畴的集成系统,解决因为各系统的差异性和信息资源的分散性所造成的检索困难。

②语义网络(Semantic Network):语义网络把概念进行分型或分类,构建概念之间的相互关系,并提供相关信息的获取。例如,查找某病毒的概念时,不但可以获取该病毒的概念和信息,还可以找到该病毒可能引起的疾病或综合征的相关概念和信息。

③信息资源图(Information Source Picture):是各种生物医学数据库的信息资源集合图,图中描述了各信息资源的范围、定位、词汇、句法和访问条件。其信息资源既可以供人类阅读,也可以被机器处理。

④专家词典(Specialist Lexicon):是为超级叙词表中的许多术语提供各种构成词的句法信息,也包括没有出现在超级叙词表中的英语单词,如动词。

(2) 关键词法。关键词(Keyword)是指出现在文献的题名、摘要或全文中,能够反映文献主题内容的专业名词或术语。关键词直接取自原文,不作规范化处理,可以提供更多的检索入口,适合计算机系统自动编制索引的需要。但由于词语没有规范化,不能进行选择和控制,容易造成漏检和误检。